Decoupled DIMM: Building High-Bandwidth Memory System Using Low Speed DRAM Devices

Hongzhong Zheng[1], Jiang Lin[3], Zhao Zhang[2], and Zhichun Zhu[1]

[1]Department of ECE
University of Illinois at Chicago

[2]Department of ECE
Iowa State University

[3]Austin Research Lab
IBM Corp.

Outline

- Challenges in DRAM memory system designs
  - Bandwidth, capacity, thermal and power
- Motivation and background
- Decoupled DRAM architecture
  - Memory performance, cost, and/or power optimization
- Experimental methodology
- Result analysis
- Conclusion

Challenges in DRAM memory system designs

- Multi-core processors

  → Increasing demands on memory's

  - **Bandwidth**
  - **Capacity**

  → **Power and Thermal**

- Advancements on memory systems
  - DDR/DDR2/DDR3, Rambus XDR
  - FB-DIMM, MetaRAM, Registered DIMM

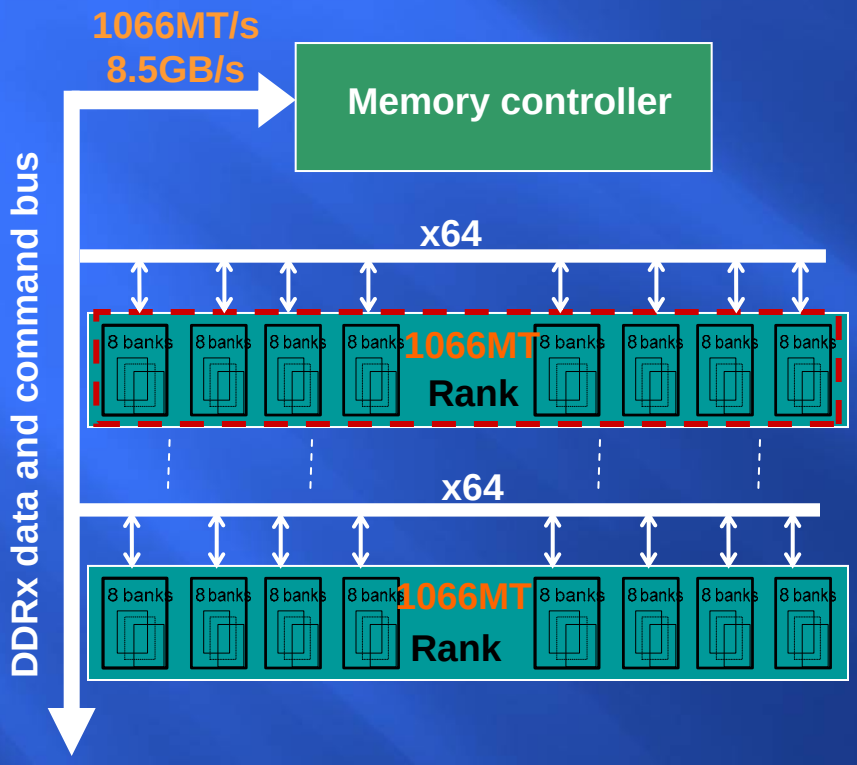
Memory Channel Design Challenges

	Channel	DRAM					Example	
	BW/CH (GB/s)	Device (MT/s)	4GB-x4-DR (W)	4GB-x4-DR ($)	I/O Road map (1Gb)	I/O Road map (4Gb)	Total Power (W)	Total Cost ($)
DDR2-667	5.3	667	10.8	83	2004	2005	65	498
DDR2-800	6.4	800	12.9	109	2006	2007	78	654
DDR3-800	6.4	800	8.0	133	2007	2008	48	800
DDR3-1066	8.5	1066	9.9	180	2008	2009	59	1080
DDR3-1333	10.6	1333	11	243	2009	2010	66	1458
DDR3-1600	12.8	1600	N/A	N/A	2010	2011	3-Channel, 24GB Xeon 2.66GHz:$1000	
DDR3-2133	17	N/A	N/A	N/A	2012	2013		

Kingston 4GB registered ECC DIMM; Power based on 2Gbit-x4 Micron device, 80% channel utilization

- Expensive for building high bandwidth channel
  - High bandwidth channel ➔ Costly and high power
  - High density DRAM device ➔ Costly and late
- Limited by DRAM device technology
  - Channel bandwidth evolvement ≤ DRAM device evolvement

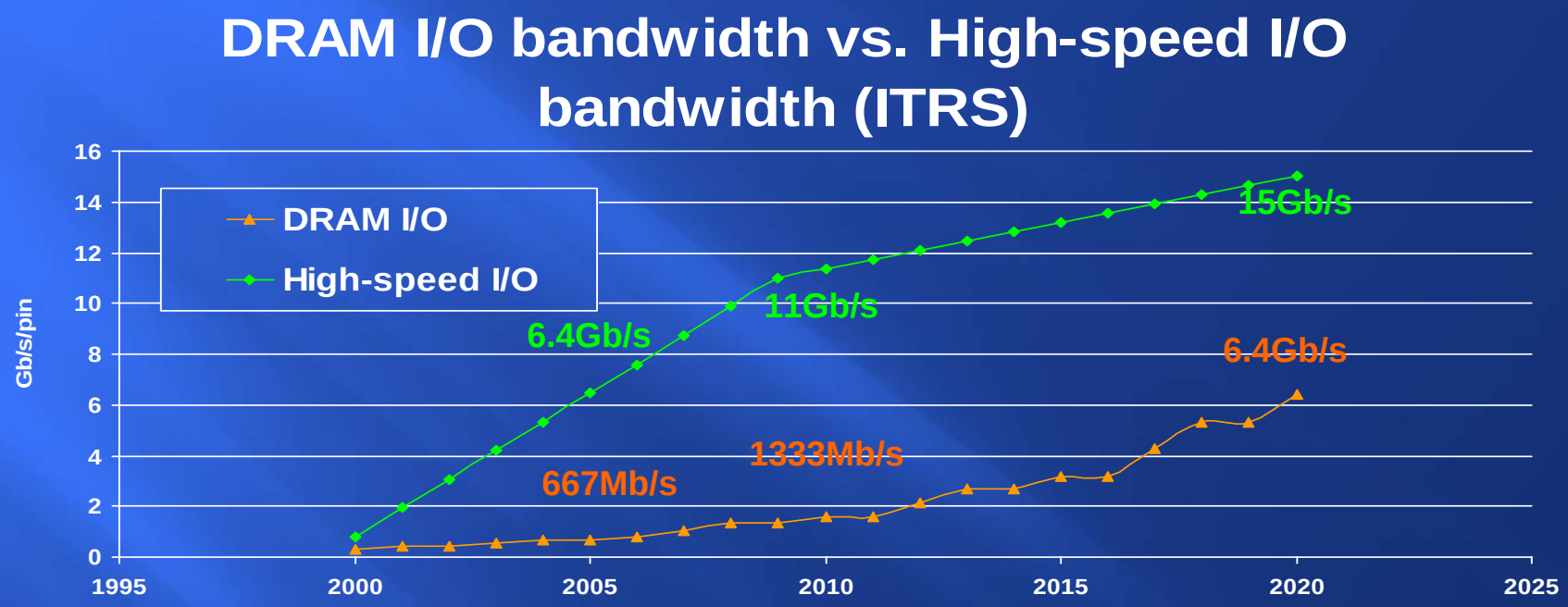
Conventional Memory Channel Organization



- Channel speed bind with DRAM devices speed
  - Rank BW = Channel BW
  - Not necessary when multi-rank per channel
- Multiple ranks per channel
  - ∑Ranks BW > Channel BW
  - NOT fully utilize the DRAM devices
  - Bandwidth bottleneck: Channel

∑Ranks BW (34GB/s) > Channel BW (8.5GB/s)

High Speed I/O Technology Available



- ↑High speed I/O > ↑DRAM speed
  - Slow evolvement of DRAM speed → bottleneck for building high bandwidth memory channel
- DRAM is optimized for capacity and cost, NOT for speed

Decoupled DIMM

- High bandwidth Channel + Low speed DRAM device ?
  - Memory channel design without DRAM evolving bottleneck
  - Benefits on performance, cost and/or power efficiency
- Design considerations
  - No changes to DRAM devices
- **Decoupled DIMM**
  - **Adding a bridge chip (Synchronization Buffer) to each DIMM/Rank**

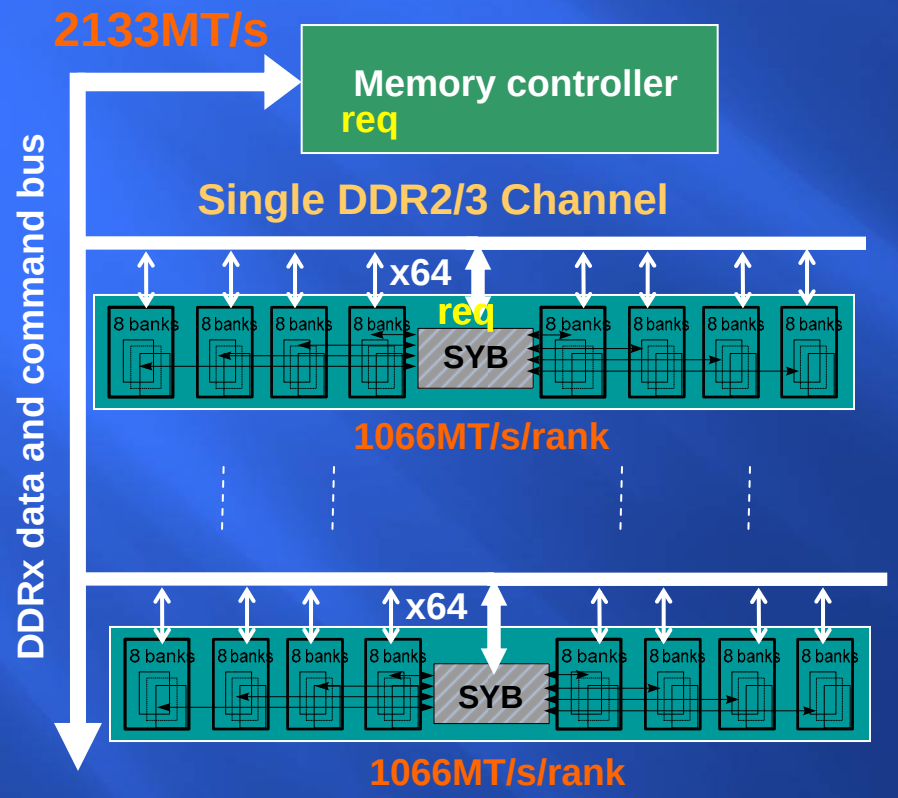
    →Breaking unnecessary bandwidth matching

    →Separating two clock domains: Channel vs. DRAM

    →Decoupling DRAM I/O tech. with Channel I/O tech.

Decoupled DIMM Design

Building high bandwidth channel using low-speed DRAM devices



Channel BW > Rank BW

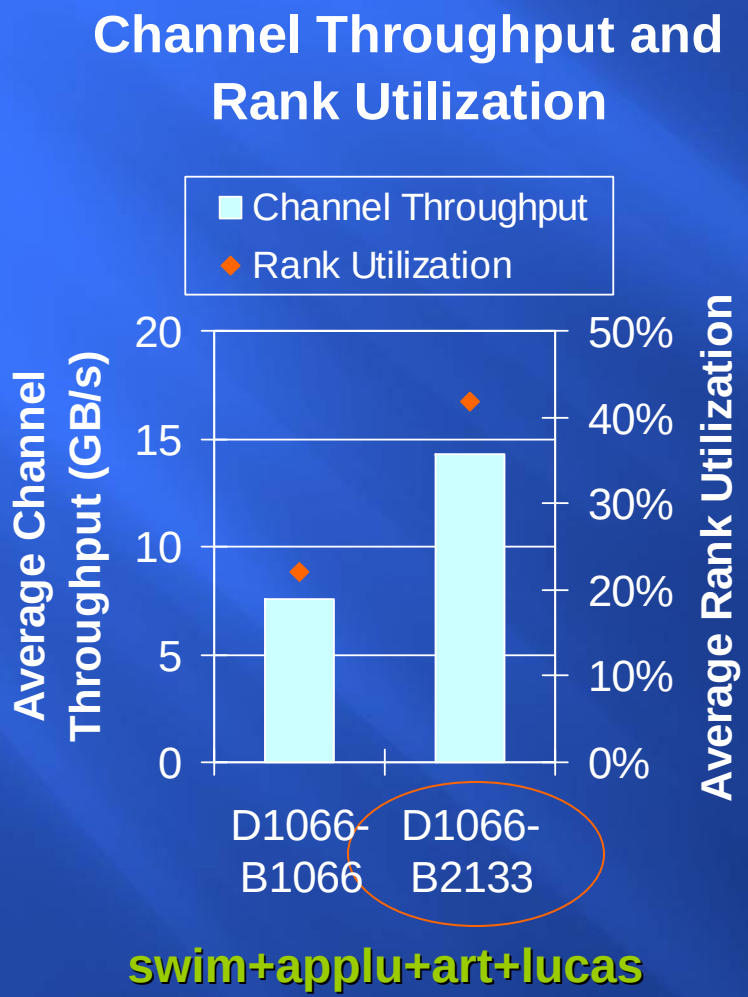
- Synchronization buffer (SYB)
  - Separating two clock domains
  - Buffering data and command
  - Introducing small latency penalty
- Breaking BW matching
  - **Channel BW > Rank BW**

    DDR3-**1066** devices → **2133**MT/s/channel
- DRAM Freq. : Channel Freq.

  1:m → 1:2, 1:3

  n:m → 2:3, 3:5

Significantly Increasing Memory Throughput

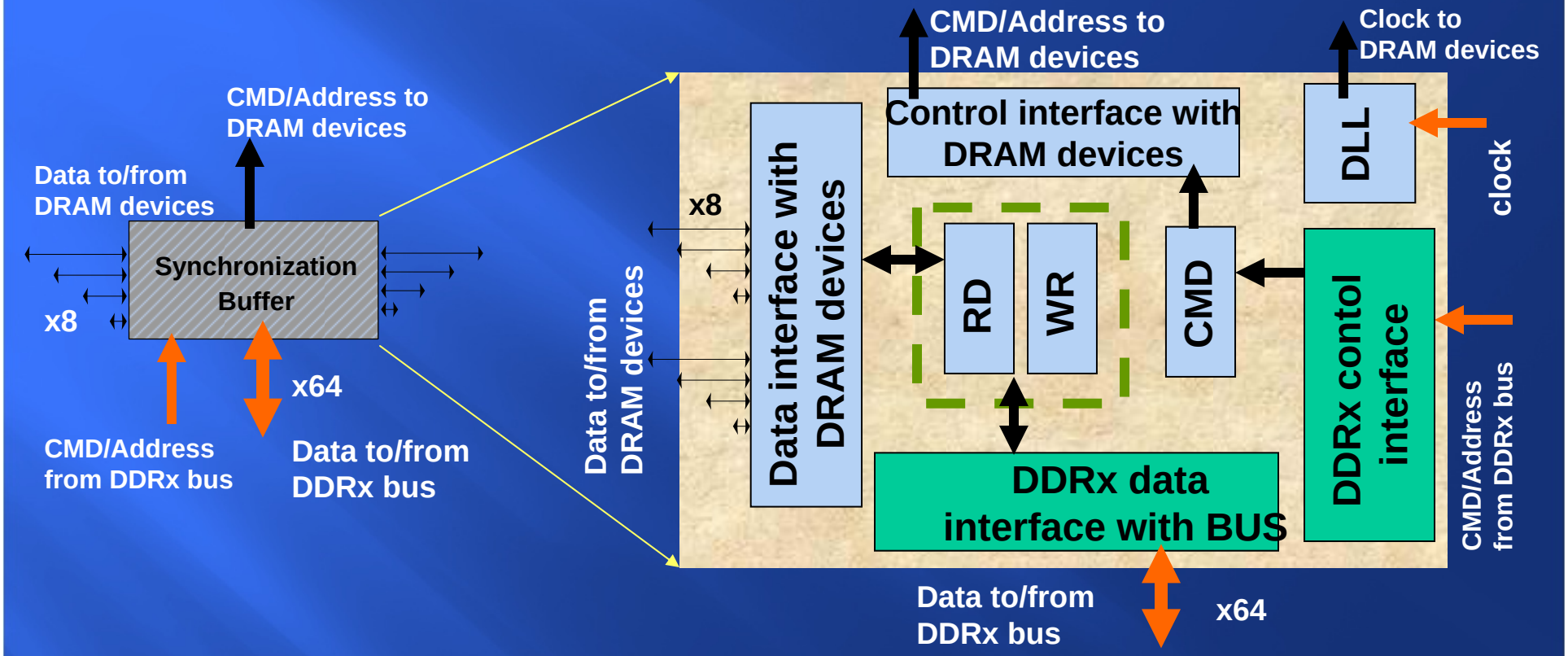


- Example:
  2CH-2D-2R, DDR3-1066, Channel 1066MT/s vs. Channel 2133MT/s
- Significantly improving memory throughput
  2 x Channel BW →
  ↑88% throughput (6.7GB/s)
- Increasing ranks utilization
  22% (1066MT/s/CH) →
  41% (2132MT/s/CH)

Benefits: Building high bandwidth channel using low-speed DRAM devices

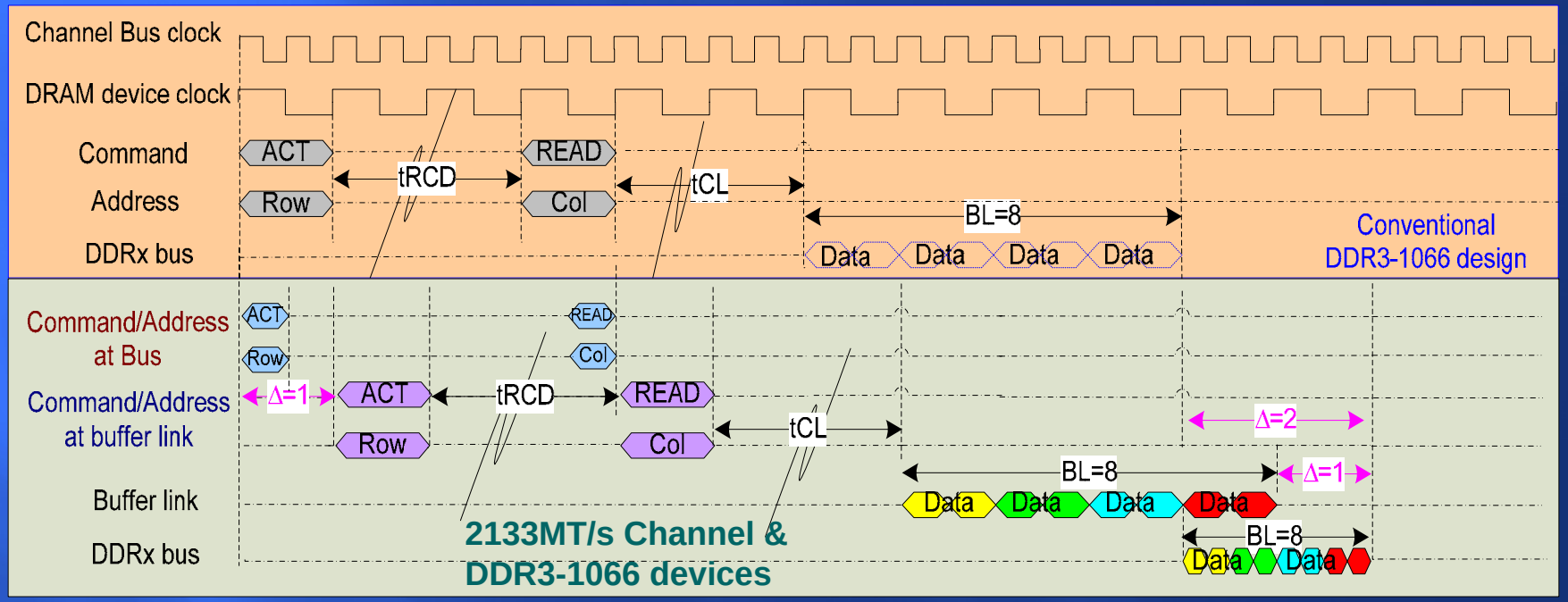
- High performance with high bandwidth
  - Channel BW > DRAM BW
- Low cost and high density
  - Low-speed DRAM devices → Low cost and high density → High BW channel
- Power/energy efficiency
  - Operating DRAM at low speed but keeping high channel BW
- More DIMMs per channel
  - Reducing electrical load of each DIMM by buffering CMD/data
- Good Reliability
  - Using standard voltage supply → High BW channel

Synchronization Buffer Design



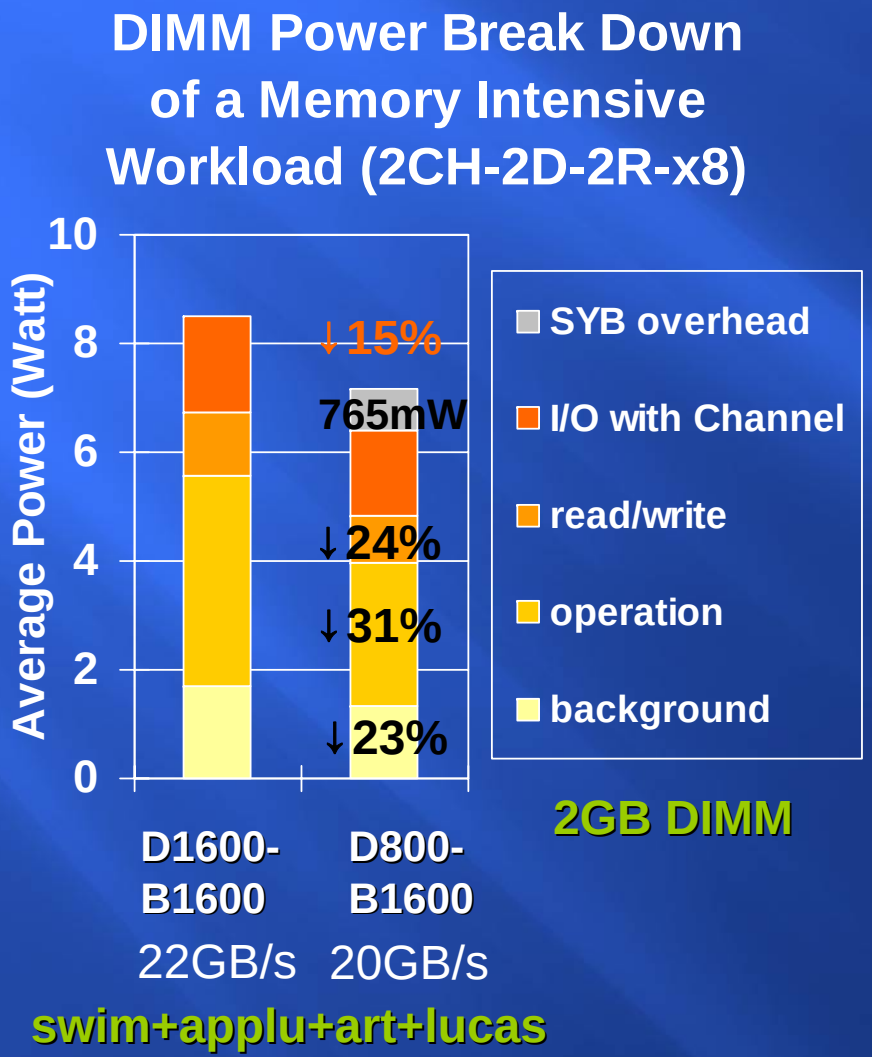
- DDRx data interface with bus
- DDRx control interface
- Delay/Phase Loop Lock
- Data interface with DRAM devices
- Control interface with DRAM devices
- Data/CMD entries inside SYB

Memory Access Scheduling



- Two level bus with SYB extends the data transfer time
  - SYB relays command and data
  - For example, DRAM devices : Channel = 1 : 2 ➔
    2 device cycles latency penalty = 1 cycle CMD delay +
    1 cycle data delay

Power Saving of Decoupled DIMM with Given Channel Bandwidth



- **Background**
  - related to power state transition and power management policies
- **Operation**
  - Activation + Precharge
- **Read/write**
- **I/O power**
  - Driving output + termination
- **SYB Overhead**

Energy Saving by Decoupled DIMM

	1600MT/s Channel & DDR3-1600	1600MT/s Channel & DDR3-800	Comments
BW (MB/s/channel)	12800	12800	Same Channel BW
Devices Freq. (MHz)	800	400	DRAM devices operating at low speed
T_{pre}, T_{act}, T_{col} (ns)	13.75	15	Small change on operation delay
Operating Cur. (mA)	120	90	25% power reduction on each operation
Background: Active Standby Cur. (mA)	65	50	>23% power reduction on background, applied most of time
T_{bl} Data burst Time (ns)	5	10	2 x data burst time by low speed devices
Read/Write Cur. (mA)	250	130	Nearly half of read/write power
SYB Latency Overhead (ns)	0	2.5	SYB latency overhead for one more I/O
SYB Power Overhead (mW)	0	382/rank	SYB power overhead for one more I/O

- Operation energy saving
  - 25% power reduction + slight change on operation delay
- Background energy saving
  - >23% power reduction + most of time

Experimental Methodology

- M5 + detailed memory performance and power simulator
- Multi-programming workloads formed by SPEC CPU2000
- Power model based on Micron power calculator
- Power management policy
  - Transiting to low power mode when no pending requests on the rank after 7.5ns
  - CC-Slow: Cache line interleaving, close page mode, and with precharge power-down slow low power mode (128mWatt, 11.25ns exit latency)
  - PO-Fast: Page interleaving, open page mode, and with active power-down low power mode (578mWatt, 7.5ns exit latency)

Major Simulation Parameters

Parameters	Values
Processor	4 cores, 3.2 GHz, 4-issue per core, 16-stage pipeline
Functional units	4 IntALU, 2 IntMult, 2 FPALU, 1 FPMult
IQ, ROB and LSQ size	IQ 64, ROB 196, LQ 32, SQ 32
Physical register num	228 Int, 228 FP
Branch predictor	Hybrid, 8k global + 2K local, 16-entry RAS, 4K-entry and 4-way BTB
L1 caches (per core)	64KB Inst/64KB Data, 2-way, 64B line, hit latency: 1-cycle Inst / 3-cycle Data
L2 cache (shared)	4MB, 4-way, 64B line, 15-cycle hit latency
MSHR entries	Inst:8, Data:32, L2:64
Memory	4/2/1 channels, 2-DIMMs/channel, 2-ranks/DIMM, 8-banks/rank, 1GB/rank
Memory controller	128-entry buffer, 15ns overhead
DDR3 channel bandwidth	800/1066/1333/1600 MT/s (Mega Transfer/s), 8byte/channel
DDR3 DRAM latency	DDR3-800: 6-6-6, DDR3-1066: 8-8-8, DDR3-1333: 10-10-10, DDR3-1600: 11-11-11

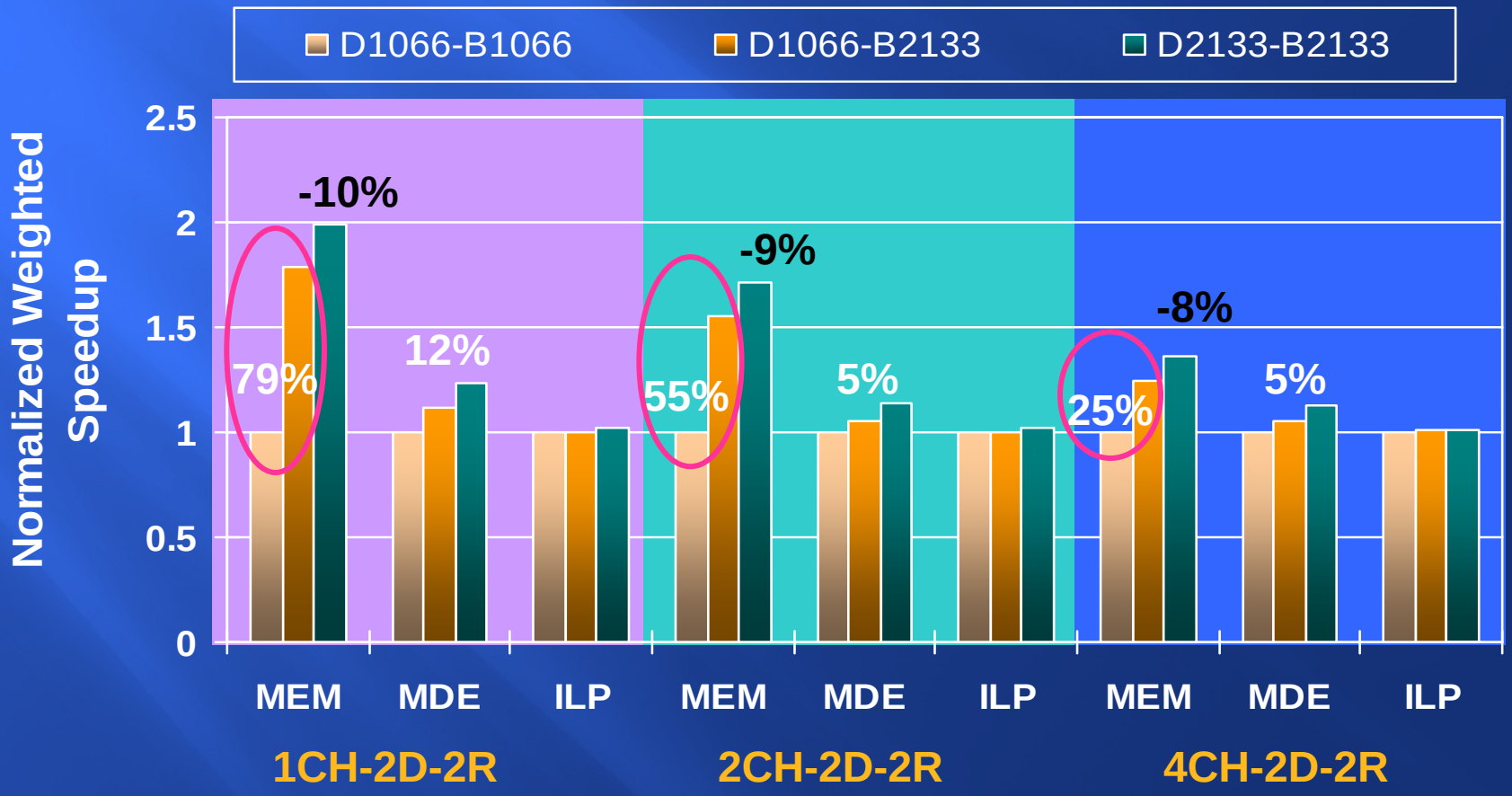
Workloads

Workload	Applications
MEM-1	swim,applu,art,lucas
MEM-2	fma3d,mgrid,galgel,equake
MEM-3	swim,applu,galgel,equake
MEM-4	art,lucas,mgrid,fma3d
MDE-1	ammp,gap,wupwise,vpr
MDE-2	mcf,parser,twolf,facerec
MDE-3	apsi,bzip2,ammp,gap
MDE-4	wupwise,vpr,mcf,parser
ILP-1	vortex,gcc,sixtrack,mesa
ILP-2	perlbmk,crafty,gzip,eon
ILP-3	vortex,gcc,gzip,eon
ILP-4	sixtrack,mesa,perlbmk,crafty

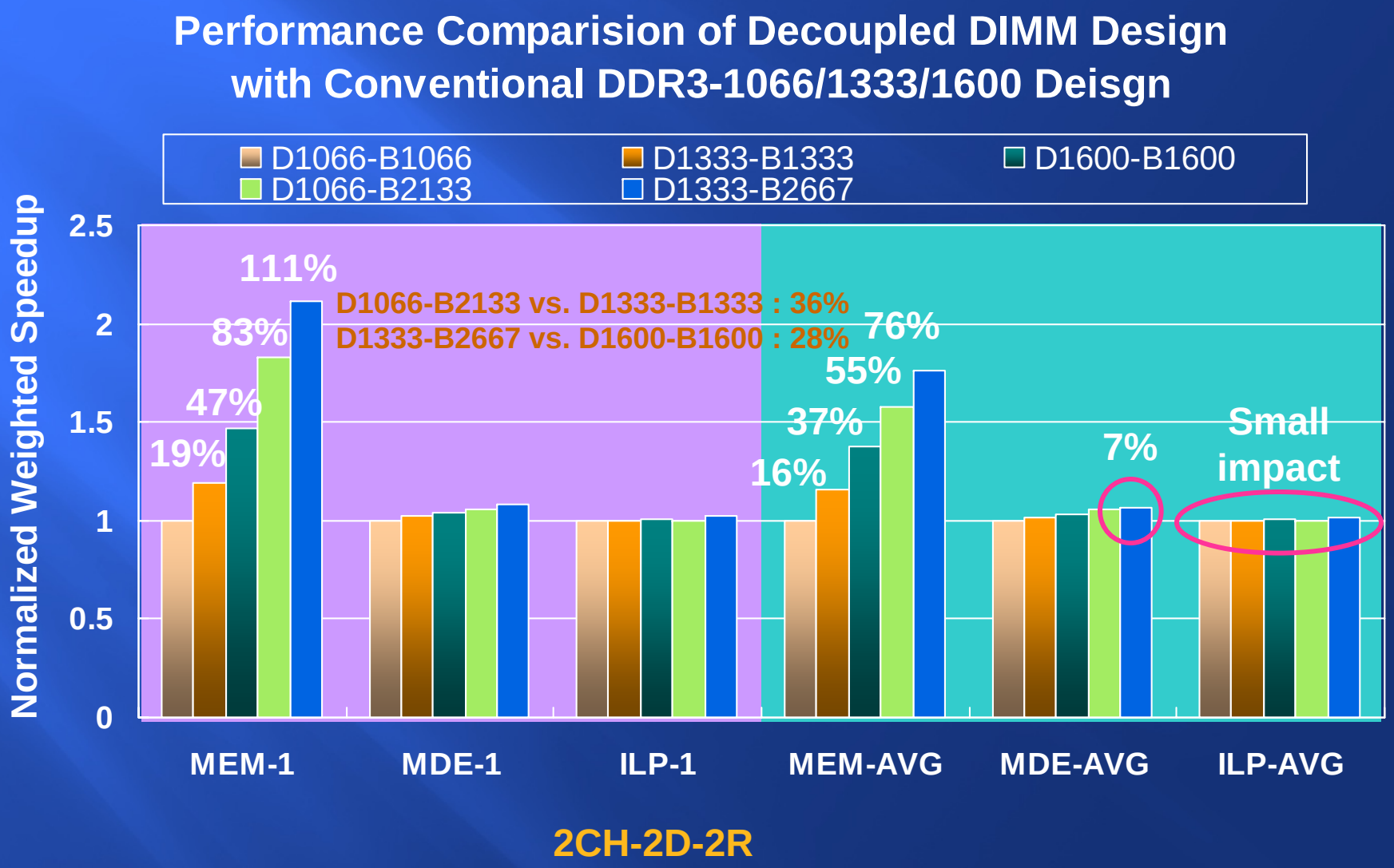
- Multiprogramming workloads randomly selected from SPEC 2000
  - MEM (memory-intensive)
  - MDE (moderate)
  - ILP (compute-intensive)
- Simulation points are picked up by SimPoint
- Performance metrics
  - Weighted Speedup
  - Harmonic mean of normalized IPCs

Average Performance of Decoupled DIMM with Given DRAM Device

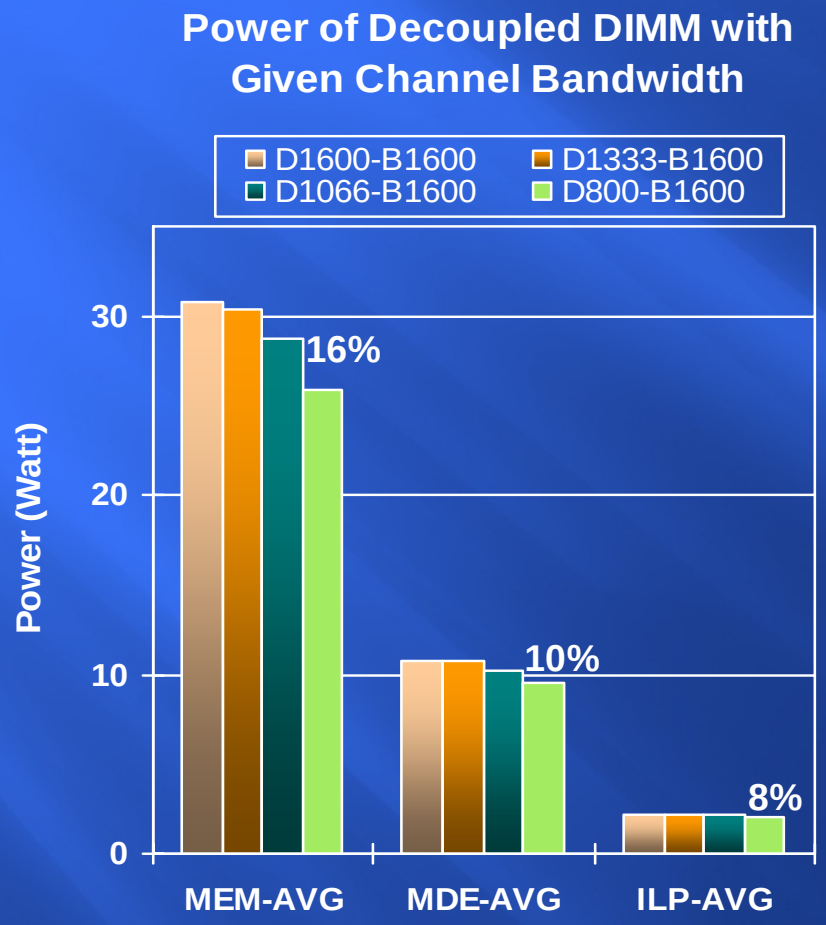
Average Performance Impact of Decoupled DIMM with Different Memory Configurations



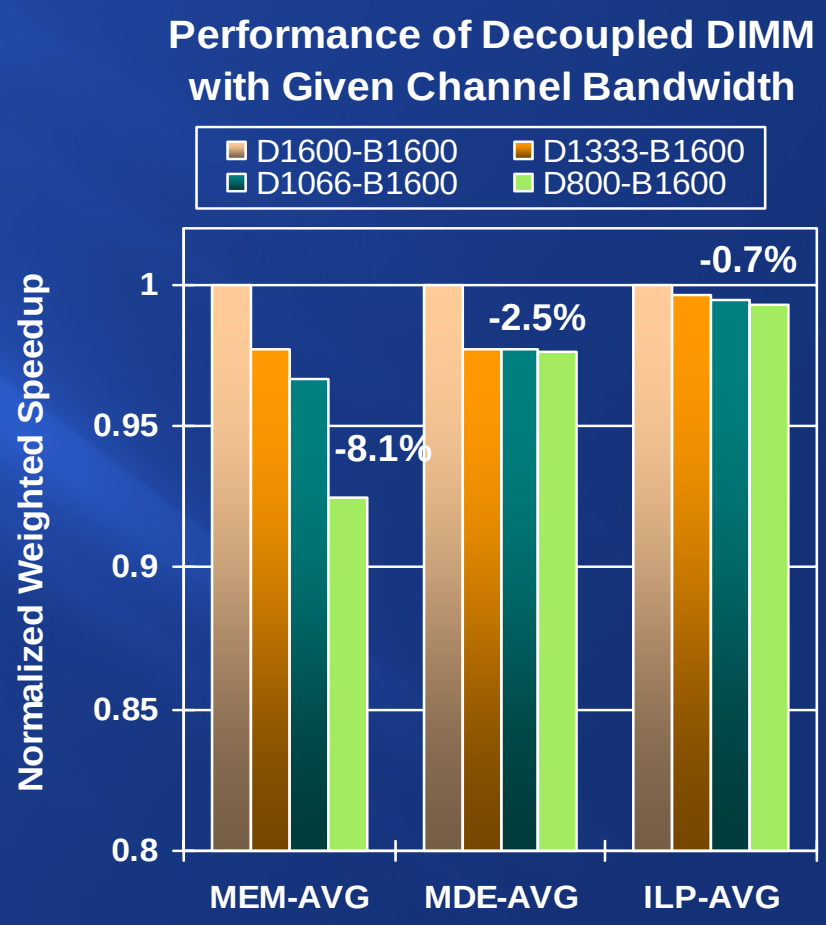
Trade-offs of Decoupled DIMM Design



Power and Performance Impact with Given Channel Bandwidth

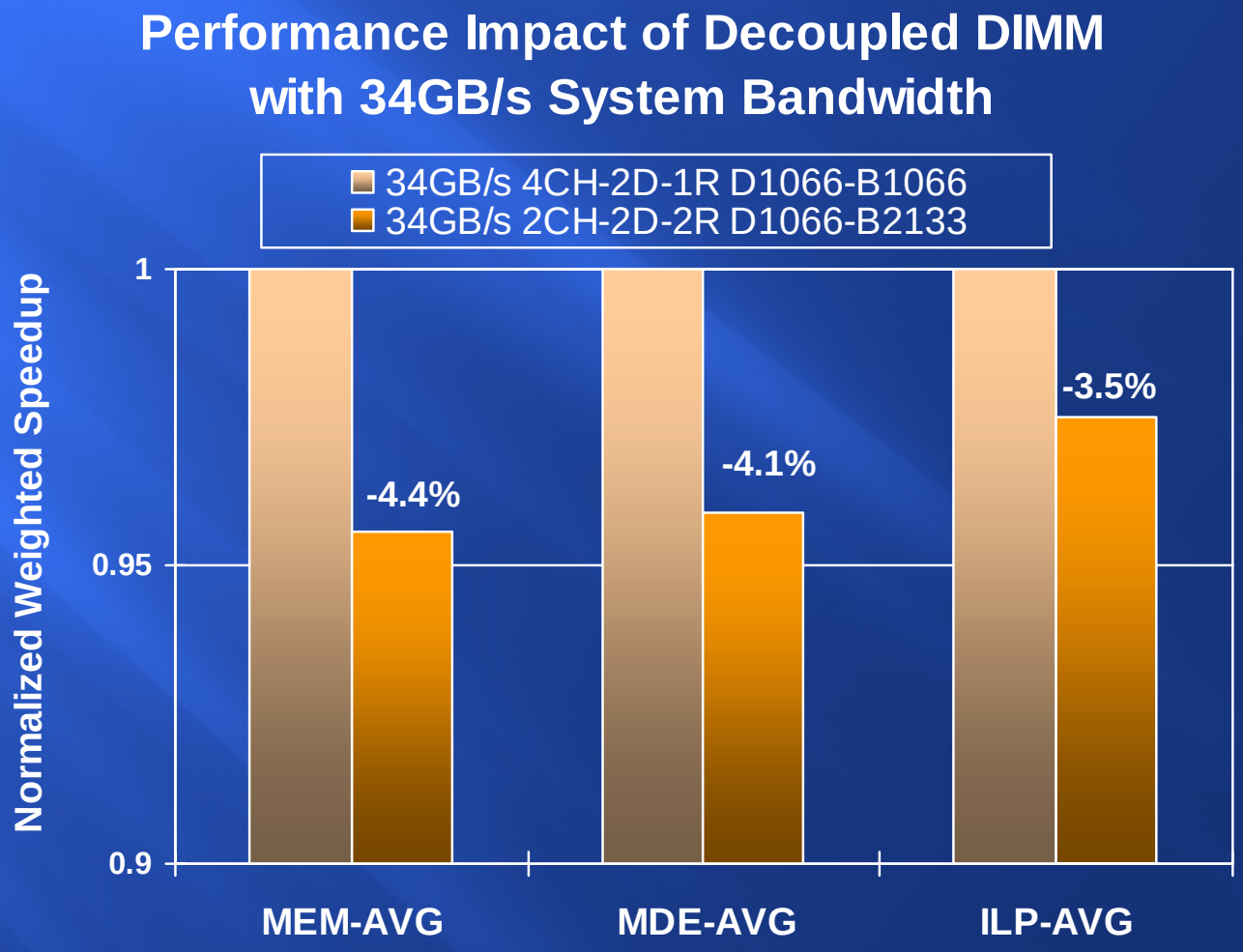


2CH-2D-2R



2CH-2D-2R

Performance Impact with Given System Bandwidth



Related Works of Decoupled DIMM

- Novel memory architecture --- Most related work
  - Mini-Rank [Zheng:MICRO2008], Threaded Memory Module [Ware:ICCD2006], Fully-Buffered DIMM [Intel2005], Register DIMM, MetaRAM [http://www.metaram.com]
- Memory system performance evaluation and analysis
  - DRAM/RAMBUS [Burger:ISCA1996, Cuppu:ISCA1999, Cuppu:ISCA2001], FBD [Ganesh:HPCA2007]
- Memory access scheduling for performance and fairness
  - Memory access reordering [McKee:HPCA1995, Rixner:ISCA2000, Hur:MICRO2004, Zhu:HPCA2005, Nesbit:MICRO2006, Mutlu:MICRO2007, Mutlu:ISCA2008, Ipek:ISCA2008]
- DRAM Low power modes optimizations.
  - Low power mode management for optimizing background power [Lebeck:ASPLOS2000, Delaluz:HPCA2001, Fan:ISLPED2001, Delaluz:DAC2002, Huang:USENIX2003, Li:ASPLOS2004, Zhou:ASPLOS2004, Pandey:HPCA2006]

Decoupled DIMM Summary

- Cost effective high bandwidth memory system design
  - Using low-speed DRAM devices building high bandwidth memory channel
- Significant benefits on performance, cost and power efficiency
  - Given DRAM devices → high bandwidth channel
  - Given channel bandwidth → power/energy saving
  - Given system bandwidth → cost effectiveness with few channels
- Small changes
  - Synchronization Buffer on DIMM
    DRAM devices design untouched
  - Small changes on memory requests scheduling